Round and Round the Garden?

University of Edinburgh
School of Law

Research Paper Series

No 2014/06

Round and Round the Garden? Big Data, Small Government and the Balance of Power in the Information Age

Dr Judith Rauhofer
Lecturer in Information Technology Law
University of Edinburgh, Law School
judith.rauhofer@ed.ac.uk

Abstract

With personal data caught in a revolving door between private and public sector access, the privacy harms arising from the monitoring of individuals are more difficult to qualify than ever. Concepts of personal data that depend on identifiability permit practices where governments and companies can single out otherwise unidentified persons on the basis of their behaviour or interests. Concepts of harm that rely on evidence of material damage ignore the way in which access to data not only maintains but re-enforces existing power imbalances. This article will look at the notion of privacy harms from an EU perspective taking into account the discussions on the role of personal data in the context of the ongoing revision of the EU data protection framework

Keywords

Information privacy, privacy harms, personal data, data protection, Big Data, power

ROUND AND ROUND THE GARDEN? BIG DATA, SMALL GOVERNMENT AND THE BALANCE OF POWER IN THE INFORMATION AGE

Judith Rauhofer

Lecturer in IT Law, University of Edinburgh, Scottish Centre for Research in Intellectual Property and Technology Law
Old College, South Bridge, Edinburgh EH8 9YL, UK
judith.rauhofer@ed.ac.uk; http://www.law.ed.ac.uk

1. Introduction

We live in an age of information, where the efficiency, competitiveness and affordability of services provided by both public and commercial bodies increasingly depends on providers' ability to process vast quantities of personal data. From government departments to small high street businesses, personal data has evolved into a currency that is tradable for goods, services, information and social contacts. This has made it vulnerable to political and economic pressures where an individual may be tempted or even forced to disclose or share their data with others in order to be able to participate in commercial, social or political life. This may result in a situation where existing power imbalances within society could be further exacerbated through an increasing asymmetry between those with access to information about others and those whose information is being accessed.

The EU data protection framework is designed to ensure that individuals (data subjects) retain overall control of their data in order to protect their right to information privacy. This right is acknowledged as a human right in Articles 7 and 8 of the EU's Charter of Fundamental Rights[1] and in the constitutions of several EU member states. However, the advent of the internet and the worldwide web has put increasing pressure on the rules comprising that EU framework. In January 2012, the European Commission therefore proposed a draft Regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)[2] with the intention of updating and harmonising the existing legal framework. This article will look at the evolving concept of personal data within EU data protection law and the way in which it continues to link protection to the identifiability of the data subject. It will also explore the prevailing understanding of privacy harms as material, quantifiable and individualized damage to data subjects' interests. It will highlight a number of "invisible harms" that have not yet been adequately addressed by the existing or the proposed new framework and will argue that those harms as well as a wider concept of personal data must be considered in the future to prevent a division of our societies into privacy haves and have-nots.

[1] 18 December 2000, OJ C364/1.
[2] 25 January 2012, COM(2012) 11.

2. The concept of personal data within the EU data protection framework

Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data Protection Directive)[3] defines personal data as "any information relating to an identified or identifiable natural person"[4]. As a rule, personal data includes personal, family and lifestyle details as well as information about a person's education, health, employment and finances.

2.1. Identifiability

A person is identifiable, if they can be identified directly through data held by a data controller or indirectly through such data in combination with data held by others[5] provided the data held by others is available to the data controller by means he is reasonably likely to use to identify the individual in question[6]. The obligations imposed on a data controller under the EU data protection framework only apply to the processing of personal data[7]. With certain exceptions[8], the processing of data "rendered anonymous in such a way that the data subject is no longer identifiable"[9] falls out with the scope of EU data protection law.

2.2. Pseudonymous data

Under the existing framework, pseudonymous data, that is, data that does not allow a data controller directly to identify an individual but which does allow him to single that individual out, is currently still considered personal data under the concept of identifiability. However, there are moves under way in the context of the ongoing reform of the EU data protection regime to apply a lower standard of protection to pseudonymous data compared to data that directly identifies the individual.

In a compromise text put forward by the European Parliament's Civil Liberties, Justice and Home Affairs Committee (LIBE)[10], the Committee suggests an amendment to the so-called "legitimate interest" conditions now included in Article 6(1)(f) of the draft Data Protection Regulation. As a general rule, this provision allows data controllers to process personal data if this "is necessary for the purposes of the legitimate interests pursued by the controller or, in case of disclosure, by the third party to whom the data is disclosed, and which meet *the reasonable expectations of the data subject based on his or her relationship with the controller*[11]". The controller's legitimate interest

[3] OJ L 281, p. 31-50.
[4] Article 2(a), Data Protection Directive.
[5] ibid.
[6] Recital 26, Data Protection Directive.
[7] Article 3(1), Data Protection Directive.
[8] For example, Article 5(3) of Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (OJ L 201, p. 37-47) restricts the use of cookies without the internet user's informed consent. This is regardless of whether the information collected by the cookie in question can ultimately be linked to an identifiable individual.
[9] Recital 26, Data Protection Directive.
[10] Report on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)), Committee on Civil Liberties, Justice and Home Affairs' Rapporteur: Jan Philipp Albrecht, Brussels, 21 November 2013.
[11] Emphasis by the author.

justifies such processing unless those interests "are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data"[12].

The LIBE Committee suggests that with regard to pseudonymous data Recital 38 should be amended to include a presumption that processing limited to pseudonymous data meets the reasonable expectations of the data subject based on his or her relationship with the controller. In addition, a proposed new Recital 58(a) states that "[p]rofiling based solely on the processing of pseudonymous data should be presumed "not to significantly affect the interests, rights or freedoms of the data subject". In practice, this could mean that the profiling of individuals without their consent could be lawful if it is based on, for example, information obtained through tracking their online behavior, personal preferences or geographical location provided that, and for as long as, the profiles created in this way are not "identified" by linking them to the name of, or other information identifying, a living individual.

The Industry, Research and Energy Committee's (ITRE) and the Committee on the Internal Market and Consumer Protection (IMCO), in their opinions on the draft Regulation[13], go even further by suggesting that the processing of personal data should - as a general rule - be lawful if only pseudonymised data are processed[14]. To this end, the ITRE opinion suggests the inclusion of an additional legal ground, which would allow the processing of pseudonymised data to safeguard the legitimate interests pursued by a controller "except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child"[15]. Both Committees justify their amendments by a need to recognize different categories of personal data and to afford them different treatment. The IMCO Committee also argues that the use of pseudonymous data will encourage good business practice safeguarding the interests of data subjects[16].

"Ensuring that personal data cannot be attributed to a data subject (since it cannot be related back to a data subject without use of additional data) helps to further promote business use of data while providing a high level of consumer protection."[17]

Similarly, the LIBE rapporteur, Jan Albrecht, in his explanatory statement, clarifies that he sees the relevant amendments as a way to encourage the anonymous and pseudonymous use of services. Data controllers should be rewarded for the use of pseudonymous data through "alleviations with regard to [their] obligations"[18].

---

[12] Article 6(1)(f), Data Protection Regulation.

[13] Opinion of the Committee on the Internal Market and Consumer Protection on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)); Rapporteur: Lara Comi, 21 January 2013, Amendment 75; and Opinion of the Industry, Research and Energy Committee on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)); Rapporteur: Seán Kelly, 23 February 2013, Amendment 101.

[14] However, it should be noted that In the case of the ITRE opinion, the new legal ground would only apply "where the data is adequately protected", see ITRE opinion, FN13, Amendment 101.

[15] ibid.

[16] IMCO opinion, FN13, Amendment 107.

[17] ibid.

[18] LIBE report, FN10, Explanatory Statement, p. 200.

2.3. Removing personally identifiable information: a discredited approach?

Of course, "faith in the privacy-protecting power of anonymisation" [19] has been severely undermined at the very least since Paul Ohm's seminal 2009 paper on the "[b]roken promises of privacy"[20]. In this paper, Ohm describes how the previous "robust anonymization assumption" that technologists and lawmakers rely on has gradually been proven to be fatally flawed as researchers began to prove "some important theoretical limits of the power of anonymization" and established what Ohm calls "the easy re-identification result"[21]. In three case studies, Ohm demonstrates the power of modern re-identification technologies, which have increasingly gained ground in the ongoing arms race between those who wish to protect individual rights through the anonymisation of personal data and those who aim to increase the utility and value of anonymous data to which they may have access through re-identification. In this context, the practical impact of re-identification on privacy laws is paradigm changing. Where

"*legislators have long relied on robust anonymization to deliver the best-of-both-worlds: the benefits of information flow and strong assurances of privacy […] [t]he failure of anonymization has exposed this reliance as misguided, and has thrown carefully balanced statutes out of equilibrium.*"[22]

Ohm argues that "at the very least, we must abandon the pervasively held idea that we can protect privacy by removing personally identifiable information (PII)"[23]. Instead, he suggests a number of approaches to deal with the privacy threats posed by what he calls the "database of ruin".

According to Ohm, this database – which he describes as "the worldwide collection of all of the facts held by third parties that can be used to cause privacy-related harm to almost every member of society"[24] - exists only in the potential. Where an entity intending to cause harm to an individual's privacy (adversary) is able to link two or more databases that may include information about that individual, then, if even one piece of data contained in one of those databases has been linked to a person's real identity, all other data that can be linked to the first piece of data, will be de-anonymised as well[25]. An example for this kind of pervasive re-identification is the way in which companies like Google, which provide different types of online services, could use identifiable information they hold in the context of one service (for example, subscriber data from a Gmail, Google + or YouTube account) to discover that user's, supposedly anonymous, search history. Google achieves this through a diverse range of tracking techniques including browser fingerprinting, IP tracking and the use of ID cookies that are placed on an internet user's device through almost every website that is linked to its Adsense programme.

The more information, identifiable or anonymised, is available to a particular adversary, the greater the risk of privacy harm to the individual. In order to restore the balance to privacy in a post-PII world, Ohm therefore suggests that regulators should shift their focus to improved enforcement, the

[19] *Ohm*, Broken promises of privacy: Responding to the surprising failure of anonymization, (2010) UCLA Law Review, Vol. 57, p. 1701
[20] ibid.
[21] ibid., p. 1706.
[22] ibid., p.1725.
[23] ibid., p.1735.
[24] ibid., p.1739.
[25] See *Narayanan/Shmatikov*, Robust De-Anonymization of Large Sparse Datasets, IEEE Symposium on Security and Privacy (2008), p. 111, at 9.

impact of different data handling techniques, the stricter control of information release to the general public rather than trusted parties and the quantity of data collected and processed.

Although the risk from re-identification of anonymous data cannot be denied, Ohm's findings do, of course, apply even more acutely to the category of pseudonymous data currently in the process of being created as a legal construct by the EU legislator. Given that the re-identifiability of pseudonymous data has never been in doubt – indeed, it is clear that under the proposed Regulation pseudonymous data would still be classified as "personal data" – it seems curious that the legislator should willingly concede a lower threshold for data controllers solely processing pseudonymous data.

It is likely that this can largely be explained by the silent assumption inherent in Europe's approach to data protection since the 1970s and 80s that the processing of anonymous and pseudonymous data - for as long as it stays pseudonymous - is the lesser of two evils compared to a situation where the data processed directly identifies an individual. Without truly knowing who the individual is, so the argument goes, it will be impossible for the data controller to abuse the position of power that comes from his possession of the data, as he will not be in a position to know at whom he should direct the exercise of that power. Under this model, the emphasis is therefore on providing the data controller with an incentive that makes it more desirable for him to **not** re-identify the data. In the context of the current discussions surrounding the EU data protection framework, this is done by proposing to introduce a regime where the processing of pseudonymous data confers certain benefits on the data controller, but where re-identification of that data would immediately result in the application of the full data protection "burden".

This argument may well have had some traction in a pre-internet, offline world, where power balances largely existed between public or private bodies and an identified individual and where those bodies' access to data about the individual determined, for example, whether a government service would be provided or denied, whether his behavior was permitted or prohibited, whether a contract was entered into or not, or whether one particular individual was given preferential treatment over another.

However, in a world where decisions about data subjects are now frequently taken without the need to identify them individually and where the potential long-term societal implications of the mass-processing of personal data – whether in identified, anonymous or pseudonymous form – increasingly raise important political, economic and constitutional concerns, it could be argued that we have to look differently at the concept of privacy harms.

3. The trouble with privacy harms: there's just no there there

One weakness in Ohm's argument is that while he assumes that the re-identification of data will open up the data subject to the threat of privacy harms by a potential adversary, he does not describe in sufficient detail the nature of those harms is and exactly what their effect would be. This leaves him open to challenge from those potential adversaries who will argue that even if re-identification should occur, this will not necessarily result in harm to the data subject. Provided that "harm" is defined in a sufficiently narrow and restrictive way, it would therefore be possible for those adversaries to call for stricter regulation and enforcement at the **processing** level, i.e. at the point where re-identification occurs and is proven to cause damage to the data subject, while leaving them free indiscriminately to **collect** data without too many restrictions. Thus it happens that online advertisers can argue that the mere tracking of internet users through the use of cookies should not be subject to any form of regulation ("no harm occurs") and that privacy and data

protection rules should only become engaged at the point when that behavioural data is mined and profiles of users' behavior are created.

3.1. Harm as individualised damage or distress

The problem with this approach is that relies solely on a concept of privacy harm that is both economic and individualized. Harm to the data subject will only arise if he himself suffers verifiable (financial) damage or distress.

Existing EU data protection law reflects this approach at least with regard to the enforceability of data protection breaches by data subjects and (in some cases) regulators. Under the Data Protection Directive, member states must provide that any person who has "suffered damage as a result of an unlawful processing operation" or of any breach of national data protection laws implementing the Directive "is entitled to receive compensation from the controller for the damage suffered"[26]. In the UK, this provision has been implemented through section 13 of the Data Protection Act 1998 (DPA).

Section 13(1) provides that an individual who suffers damage because a data controller breaches any of the provisions of the Act is entitled to compensation from the data controller for that damage. Section 13(2) includes a similar cause of action where the breach has caused the data subject to suffer distress. However, section 13(2)(a) makes it clear that data subjects can only be awarded compensation for distress if they can show that they have also suffered financial loss. Similarly, the UK Information Commissioner has the right to serve a data controller with a monetary penalty notice of up to £500,000 if he is satisfied that there has been a serious contravention of one of the data protection principles, the contravention was of a kind likely to cause substantial damage or substantial distress, and the data controller had the requisite *mens rea*[27].

While the damages provisions in section 13 of the DPA have attracted widespread criticism[28], the need for the existence of **either** individualized financial loss **or** distress is not generally questioned. At the same time, the absence of significant case law where data subjects have brought successful claims for damages under section 13 demonstrates the difficulties data subjects face when to rely on that cause of action in court. On the contrary, in the case of *Douglas & Ors v Hello! Ltd & Ors*, which concerns the unauthorized publication of photos of Michael Douglas and Catherine Zeta Jones' wedding by Hello! Magazine, the High Court ruled that while the conditions of section 13 were certainly fulfilled, it did "not see it as adding a separate route to recovery for damage or

[26] Article 23(1), Data Protection Directive.

[27] Section 55A, DPA.

[28] Most recently from Lord Justice Leveson in his report on the culture, practices and ethics of the press, who proposed that the DPA should be amended so that compensation can be awarded regardless of pecuniary loss, see *Leveson LJ*, An inquiry into the culture, practices and ethics of the press, The Stationer Office, 29 November 2012, Part H, Chapter 5, para 2.61; available at http://www.official-documents.gov.uk/document/hc1213/hc07/0780/0780.asp; last visited on 6 January 2014. Leveson's proposal has also been welcomed by the Information Commissioner, who highlights that the European Commission have questioned whether the UK has properly implemented the Data Protection Directive in this respect, see ICO, The Information Commissioner's Response to the Leveson Report on the Culture, Practices and Ethics of the Press, 7 January 2013, p. 11; available at http://www.ico.org.uk/about_us/consultations/~/media/documents/consultation_responses/ico_response_to_leveson_report_012013.ashx, last visited on 6 January 2014. The need to have suffered financial damages before a claim on the basis of distress can be made was also questioned by Tugenhat J in *Vidal-Hall and others v Google Inc [2014] EWHC 13 (QB), 16 January 2014.*

distress beyond a nominal award [of £50]"[29]. Similarly, in the later case of *Campbell v MGN Ltd.*, the House of Lords found that a claim under the DPA "adds nothing" to a claim for breach of confidence that was part of the same action[30]. More recently, in the case of *Halliday v Creation Consumer Finance Ltd (CCF)*, the Court of Appeal gave some guidance on the way in which it assessed the element of distress. The appellant, Mr Halliday, had purchased a television set through a credit arrangement with Creation Consumer Finance Ltd (CCF). Protracted dealings and court proceedings between the parties resulted after CCF committed numerous data breaches, which included providing incorrect data concerning Mr Halliday to a credit referencing agency; that data was then made available to third parties for a period of four months. While the Court acknowledged that the appellant would have felt "frustration at these prolonged and protracted events", it found "no contemporary evidence of any manifestation of injury to feelings and distress" and consequently decided that "the sum to be awarded should be of a relatively modest nature *since it is not the intention of the legislation to produce some kind of substantial award*"[31]. Similarly, in 2013 the First-tier Tribunal (Information Rights) overturned two monetary penalty notices issued by the Information Commissioner against data controllers on the basis that their breaches of the DPA did not create a likelihood of causing substantial damage or substantial distress. In *Scottish Borders Council v The Information Commissioner*, a data processor operating on behalf of a local authority had discarded employee pension files in a supermarket recycling facility[32]. The ICO had argued that that substantial damage or substantial distress may have arisen from confidential personal data being seen by a member of the public, from the data having been potentially further disseminated (even if that did not actually happen) and from potential disclosure to untrustworthy third parties, thus creating the risk of identity fraud and possible financial loss. Although the Tribunal acknowledged that the appellant may have breached the data security principle, it held that it was unable to construct a likely chain of events which would actually lead to the harm described by the ICO. In *Christopher Niebel v The Information Commissioner*, the appellant had sent unsolicited text messages to internet users and committed other breaches of the Privacy and Electronic Communication (EC Directive) Regulations 2003[33]. While the Tribunal accepted that people might incur charges for replying "stop" to the text messages, it rejected the ICO's contention that damage would arise because of a text message filling up phone memory and that "opportunity costs" arose in the time taken to react to and delete an unwanted text message. It also found that while the messages were likely to cause widespread **irritation**, it was unlikely that they would cause data subjects any **distress**. These cases therefore show that the concepts of damage and distress do not form a sound basis for the enforcement of data protection principles and the protection of individuals from privacy harms.

Unfortunately, data subjects can currently not expect a change in legislative attitude from the draft Data Protection Regulation. Article 77 of the draft Regulation includes the wording of Article 23 of the Directive in almost unchanged form.. Although Recital 67 takes some steps better to define the circumstances in which a data protection breach might adversely affect the personal data or privacy of a data subject ("where it could result in, for example, identity theft or fraud, physical harm, significant humiliation or damage to reputation"), Recital 68 continues to refer to "damage to personal and economic interests". This approach overlooks the myriad of intangible, collective and

[29] *Douglas & Ors v Hello! Ltd & Ors [2003] EWHC 786 (Ch) (11 April 2003)*, at para. 239.
[30] *Campbell v MGN Ltd [2004] UKHL 22* (6 May 2004), at para. 130
[31] *Halliday v Creation Consumer Finance Ltd (CCF) [2013] EWCA Civ 333 (15 March 2013)*, at para. 35. Emphasis added by the author.
[32] *Scottish Borders Council v The Information Commissioner, EA/2012/0212, 21 August 2013.*
[33] *Christopher Niebel v The Information Commissioner (EA/2012/2060), 14 October 2013.*

societal harms that can be caused by a data protection breach. It also ignores the potential long-term effect of such breaches, which may result in privacy harms further down the line, which we may not yet be able fully to determine.

4. Invisible harms

Conversely, the recent debate surrounding privacy and data protection has been characterized by a subtle change in rhetoric. This debate, in which data controllers aim to persuade legislators of the merits of a "risk or harm based approach", threatens to undermine some of the most fundamental tenets of the European concept of privacy and data protection as a fundamental human right. This includes *inter alia* the concepts of purpose limitation and data minimization.

4.1. Purpose limitation

European data protection law is underpinned by the concept of purpose limitation. This means that data collected for one purpose must not be used for other incompatible purposes. The purpose limitation principle is currently set out in Article 6(1)(b) of the Data Protection Directive, which has been implemented in all EU member states. It is designed to ensure that data subjects are at all times aware who is using their data and in what context. In its 2013 Opinion on purpose limitation[34] the Article 29 Working Party highlights the importance of the purpose limitation principle for maintaining an adequate "information power balance" between the data subject, the data controller and any third parties that might be interested in further processing the personal data in question. It also advises particular caution with regard to a change in purpose if "the data subjects, or any third party on their behalf, were obliged to provide the data under law"[35] or if the data collection is based on a contractual relationship where one party is in a significantly weaker bargaining position. It emphasizes that in those cases "the balance of power between the data subject and the data controller […] should be examined"[36].

Despite such noble encouragement (or discouragement, depending on how one looks at this), we are now increasingly witnessing a situation where personal data moves freely through a "revolving door" from public to private (and *vice versa)*, public to public and private to private bodies, with little regard for the purposes for which it was originally collected. The prevailing attitude inherent in many contemporary processing practices seems to be that "if the data is already there, we should be able to use it". This was most visibly illustrated in a speech by Cecilia Malmström, European Commissioner responsible for Home Affairs, in February 2011, when she presented a proposal for an EU Passenger Name Record (PNR) Directive[37] intended to fight serious crime and terrorism. If adopted, the Directive would oblige air carriers to provide data on all passengers entering or departing from the EU to so-called national passenger information units. In her speech and in the subsequent press conference[38], Malmström went to some length pointing out that airlines would not

[34] Opinion 03/2013 on purpose limitation, WP203, 2 April 2013; available at http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf; last visited on 6 January 2014.
[35] ibid., p. 24.
[36] ibid.
[37] Directive on the use of Passenger Name Record data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime, COM(2011) 32 final, Brussels, 2.2.2011.
[38] Proposal for a directive on the use of passenger name record data in the EU:- press conference by Cecilia MALMSTRÖM, European Commissioner for Home Affairs, 2 February 2011. An audio recording of the press

be required to collect **new** data about their passengers. Instead, the obligation to transfer would solely refer to information that those airlines had already collected and stored for their own commercial purposes. Malmström reiterated this point no fewer than seven times during the half hour event, emphasizing again and again that "the information has been collected for decades"[39], that the "data is already there"[40] and that "we are not building up new data systems, we are just finding systems to use the data that is there in a more intelligent way for all these purposes and fighting serious crime and terrorism"[41]. This approach not only clearly illustrates the existence of Ohm's "database of ruin", it also highlights the invisible privacy harm that can occur when the contextual integrity of personal information disclosure is breached. From the data subjects' perspective, it is now almost inevitable that sooner or later data they disclose to one entity will be shared across two or more public or private entities without their specific consent and often without their conscious knowledge. It thus becomes impossible for the data subject to appreciate, at the time of collection, how long their data will be stored, how it will be used in the future, for what purposes and by whom. Data subjects, when disclosing their data to anyone, are thus unable to make an informed decision about the risks involved in that disclosure and they are consequently prevented from taking reasonable precautions against uses of their data, which may subsequently harm them. As Solove points out with regard to the wide array of reasons for which governments may share citizens' personal data, "[t]he potential future uses of any piece of personal information are vast, and without limits on accountability for how long that information is used, it is hard for people to assess the dangers of the data's being in [another party's] control"[42]. This represents a shift in the balance of "information power" in the data controller's favour the effects of which have yet to be determined.

4.2. Data minimisation

The principle of data minimisation has long been a cornerstone of the EU data protection framework. Article 6(1)(c) of the Data Protection Directive provides that personal data must be "adequate, relevant and **not excessive** in relation to the purposes for which they are collected and/or further processed". The draft Regulation seeks to further strengthen this concept by providing that data must "only be processed if, and as long as, the purposes could not be fulfilled by processing information that does not involve personal data"[43]. Section 3a of the German Federal Data Protection Act already includes a similar requirement.

Nevertheless, the last decade has shown a change in the attitude of data controllers both in the public and the private sector, who would ideally like to move to a concept of "all the data all the time". The recent revelations by Edward Snowden regarding the mass surveillance and indiscriminate collection of the content of, and traffic data generated by, electronic communications through the US and UK security agencies provide the most succinct evidence for this type of approach. But further proof, should it be needed, can also be found in other areas like commerce and research. The concept of "Big Data" essentially relies on the idea that nearly every commercial, public policy, research or health problem could be solved (or solved more cheaply), if only we had

---

conference (Reference 75222) can be accessed via the European Commission's Audio-visual Service at http://ec.europa.eu/avservices/audio/audioDetails.cfm?ref=75222&sitelang=fr; last visited on 6 January 2014.

[39] ibid., at 2:38.

[40] ibid., at 29:57.

[41] ibid., at 28:39.

[42] *Solove*, Nothing to Hide: The False Tradeoff between Privacy and Security, Yale university Press, New haven and London, (2011), p. 28.

[43] Article 5(1)(c), draft Data Protection Regulation.

enough data. Companies and governments alike have therefore begun to reassess the value of their databases not for the purpose of facilitating existing relationships with their customers and citizens, but with a view to extract maximum profit or utility from the data they hold on those customers and citizens.

Companies will seek to exploit their customer's data to maximize revenue, for example by tailoring services to them, targeting only the most profitable, or commercially exploiting the information they hold by selling it, in identifiable or aggregate form, to third parties. This can result in a number of harms to the individual customer that have not yet been sufficiently explored. For instance, the tailoring of service to a customer's known preferences may increase sales to the customer, but it may also trap the customer in a type of filter bubble[44], in which he is no longer exposed to different goods, services or experiences. In a pre-sale context, much of this demand-side manipulation can be done without directly identifying the customer, who is thus reduced to the picture he presents based on previous purchases, websites visited and services used. The long-term effect this could have, on a societal level, on the available opportunities for cultural, political or economic self-development still needs to be adequately explored. However, it is clear that this type of harm is neither quantifiable, nor does it affect the data subject at a solely individualized level. For example, a lack of exposure of citizens to different political views could affect their individual ability to participate in the democratic debate. However, in the long term, this could also have severe effects for the general level of democratic participation within society as a whole, thereby weakening existing governance structures. However, the "targeting" approach is rarely challenged on an individual level as it supports the notion that in a free market society the customer gets what the customer wants in the most efficient way.

There are also concerns that the profiling and subsequent targeting of individuals could lead to price discrimination and predatory marketing to particular groups of consumers[45]. The potential harm caused by this practice is often rejected on the basis that it may instead lead to a quasi-socialist approach, where each person will pay for goods and services according to their ability. However, when factoring in the existence of basic human needs and existing information and bargaining asymmetries between customers and suppliers, it is much more likely that it will lead to a situation where those at the bottom rung will be forced to pay over the odds for essential goods or services, as commercial data controllers will quickly be able to single out those who have little choice but to agree to unfavorable terms. As before, this could be done on the basis of anonymous or pseudonymous data without identifying users or customers as individuals. Social, political and commercial "exclusions zones" can therefore be operated solely on the basis of data that allows for an individual's classification as a member of a particular social, political or economic group.

Finally, the law of unintended consequences should also taken into account when advocating the use of Big Data for public policy purposes. For example, the publication of government crime statistics under the UK government's "Open Data" initiative has quickly led to the creation of UK crime maps[46], which rate neighborhoods in terms of security and types of crime committed. While supporters will argue that this should provide an incentive to security services to ensure that the crime rate in areas under their control is reduced, in practice, the granularity of the maps, which

[44] For a detailed exploration of this phenomenon see *Pariser*, The Filter Bubble: What The Internet Is Hiding From You, Penguin Books, London, (2011).

[45] For a more recent exploration of these issues, see *Newman*, The Costs of Lost Privacy: Consumer Harm and Rising Economic Inequality in the Age of Google (2013); available at SSRN: http://ssrn.com/abstract=2310146, last visited on 6 January 2014.

[46] The crime maps can be accessed by postcode via http://www.police.uk, last visited on 6 January 2014.

show, for example, the number of violent and sexual offence, burglaries or antisocial behavior that occurred in any particular street, can pose a great threat to the privacy not only of offenders but also of crime victims. In addition, it is likely that over time, certain areas are turned into "green", orange" or "red" zones, which will attract benefits or penalties for their inhabitants. For example, this could include higher or lower insurance premiums, or discounts or surcharges for the provision of certain services[47]. While some may argue that this information also allows individuals the choice to decide where they want to live on the basis of a range of criteria (safety, insurance premiums, etc.), in practice, this is an unrealistic assumption given the myriad of factors that may force an individual to move to a particular area or prevent him from moving away. Work, family or the simple fact of being trapped in negative equity following the decline of house prices in a particular neighborhood could all be named in this regard. As a result, the potential societal implications of this use of data should be examined in much greater detail before it is shared openly with a view to encouraging its further commercial exploitation.

5. Conclusion

With personal data now caught in a revolving door between private and public sector access, the privacy harms arising from the monitoring of individuals are more difficult to qualify than ever. Concepts of personal data that depend on identifiability permit practices where governments and companies can single out otherwise unidentified persons on the basis of their behaviour or interests. Concepts of harm that rely on evidence of material damage ignore invisible privacy harms that maintain or even re-enforce power imbalances that already derive from existing information asymmetries. The potential long-term impact of this shift in information power needs to be assessed in much greater detail before the decision is taken to pursue a purely "risk-based" approach.

While there will always be those who are better equipped to navigate this new information environment and who will use it to their advantage (i.e. by benefitting from information disclosure of others while protecting their own privacy) it is likely that they will be belong to the group of well resourced, well educated and well informed individuals that already hold much of the power in modern society. As computer security expert Dan Geer argues, the protection of privacy may well be a young person's problem[48]. While anyone old enough[49] or well off enough "can opt out of many of the corporate data collection schemes and live out the remainder of [their] days unaffected by what [they] might be missing out on, "[a]nyone under 40 has no such option, or at least no such easy option" if they want to continue to a socially, politically and economically active life. The question therefore is not (or not solely) whether changes to the existing EU data protection regime will be sufficient to tackle the new privacy issues raised by the online environment, but whether we must replace familiar concepts like identifiable personal data and easily quantifiable harms with new ways of thinking about privacy and data protection, if we want to prevent the growing

[47] This brings to mind the American Civil Liberties Union's famous spoof "pizza video", in which the customer of a pizza delivery service is charged a surcharge because the driver is forced to deliver to an orange zone. The video can be accessed at https://www.aclu.org/ordering-pizza, last visited on 6 January 2014.

[48] *Geer*, Tradoff in Cyber Security, Keynote delivered to the University of North Carolina Cyber Security Symposium, 9 October 2013; available at http://geer.tinho.net/geer.uncc.9x13.txt, last visited on 6 January 2014.

[49] For example, it can likely be argued that the historically lower number of older users of social media services can at least partly be explained by the fact that their social circles were already well established before the advent of social media services and are not dependent to the same extent as those of, say, teenagers, on the use of online communications tools for their everyday operation and continued existence.

information asymmetry, and the invisible harms resulting from it, that is capable of causing long-term harm to our societies.